

Artikelgranskning

Detta dokument kan användas som stöd vid en artikelgranskning, men ska ej ses som ett heltäckande dokument för vetenskapsmetodik. Vissa av rubrikerna är mer eller mindre relevanta beroende på artikelns studiedesign. Vid eventuella upptäckta felaktigheter i dokumentet, vänligen kontakta anne.ortqvist@ki.se

Författare:

Titel, tidskrift:

Datum:

Studiedesign:

Studiens syfte/hypotes:

Studiepopulation, inklusions- och exklusionskriterier:

Följsamhet, bortfall:

Exponering (undersökt riskfaktor/intervention), referensgrupp:

Utfall, uppföljningstid:

Huvudresultat (statistisk signifikans, klinisk relevans):

Intern validitet (systematiska fel):

Extern validitet (generaliserbarhet):

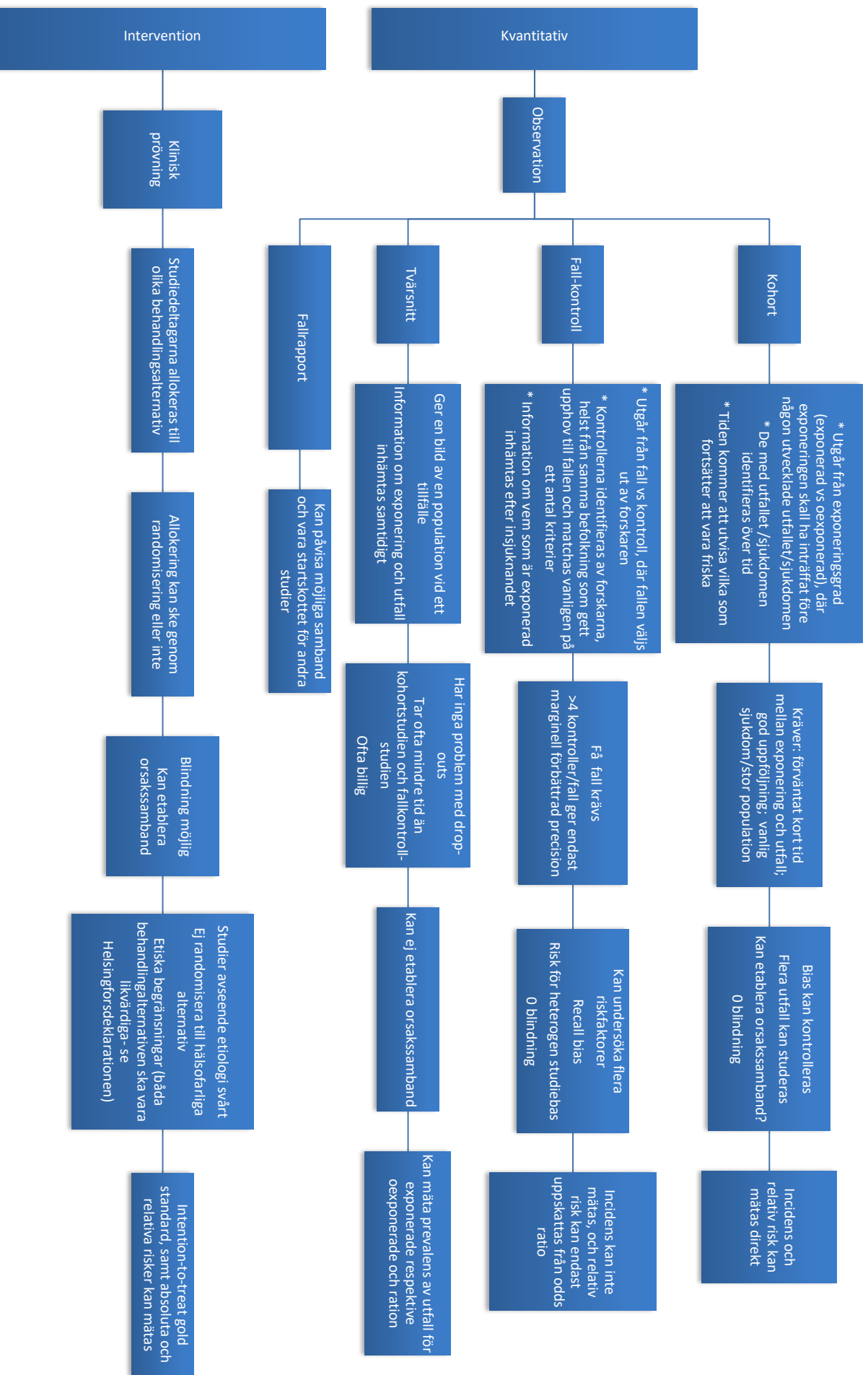
Precision:

Sample size och study power:

Vad talar för och vad talar emot denna studie?

För	Emot

Studiedesign: Vilken typ av studiedesign har använts? Kvantitativ eller kvalitativ metod? Nedan finns ett flödesschema (modifierat efter figur 11.1 i "Att börja forska", se referenser nedan) för de vanligaste kvantitativa studiedesignerna. För mer information om systematiska översikter se exempelvis SBU's metodhandbok: <https://www.sbu.se/metodbok?pub=48286>.



Studiens syfte/hypotes:

- Är studiens syfte och hypotes specifik och angiven på förhand (a priori)? Utan en specifik hypotes är det svårt för läsaren att avgöra om metodvalet är korrekt. Exempel allmän hypotes: "Antibiotika hos gravida ökar risken för astma hos barn". Exempel på specifik hypotes: "Bredspektrum-antibiotika i tredje trimestern hos kvinnor som genomgått vaginal förlossning ökar risken för allergisk astma första levnadsåret hos barnet".

Studiepopulation:

- Är studiepopulationen bestämd på förhand? Både för observationsstudien och den klinisk prövningen är inklusionskriterier och exklusionskriterier avgörande för att man ska kunna bedöma studiens externa validitet (se nedan).
- Hur har studiedeltagarna rekryterats: alla som fanns tillgängliga i ett register? Av en studieansvarig som handplockat de patienter hen tror har bäst möjlighet att genomföra studien, eller alla konsekutiva patienter? Vid ett enskilt eller flera testcenter?
- Inklusionskriterierna bör vara rimliga så att det motsvarar den "genomsnittliga patienten".
- Stämmer inklusionskriterierna med eventuella existerande internationella definitioner/diagnoskriterier som finns?
- Exklusionskriterierna bör också vara relevanta, och ej för exkluderande. Ex: Ska man exkludera gravida kvinnor från vaccinstudier? Är det rimligt att exkludera ex missbrukare från en klinisk prövning för att man inte tror att hen skall klara av att fullfölja studien? Är det rimligt att exkludera en patientkategori man tror kan vara extra känslig för biverkningar?

Följsamhet (compliance/adherence), bortfall (drop-out): Bortfall kan vara försmädligt för en klinisk prövning om de som hoppat av studien gjort det pga biverkningar, de ogillar vården de får på "studiecentret", eller att protokollet för studien är för jobbigt. Först och främst skall man motverka bortfall i så stor utsträckning det går. Ibland är det dock oundvikligt. Ett sätt att komma runt problemet med bortfall i en klinisk prövning är att använda *Intention-to-treat* (ITT) analys, där man inkluderar alla individer som randomiserats, till skillnad från *per protocol* där man endast analyserar de som följt hela behandlingsinterventionen, vilket kan introducera bias. Om en intervention är sant effektiv, kommer en ITT-analys att ge ett "unbiased" mått på effektiviteten av intervention jämnt mot följsamheten till interventionen. Vid per protocol-analys tenderas effekten att överskattas eftersom följsamheten till ordinerad behandling i praktiken inte är hundra procentig, dels finns en stor risk för bias eftersom följsamhet till behandling ofta är relaterat till kliniska utfall.

Det är också viktigt att redovisa bortfall i enkätstudier, och hur många påminnelser studiedeltagaren har fått.

Oavsett orsak för avhopp är ett mindre bortfall bättre än en stort, givet att det inte är någon relevant skillnad i mekanismerna till bortfallet. Ett bortfall bedöms ofta som litet vid <10%, medelstort 10-19%, stort 20-29% och mycket stort om >30%. Är det ett selektivt, men litet bortfall, är behovet det egentligen inte vara sämre än ett stort icke-selektivt.

- Redovisas bortfall från den aktuella studien under granskning? Finns det ett detaljerat flödesdiagram (i princip obligatoriskt i den kliniska prövningen – se hänvisning till checklistor nedan) där man redovisar antal människor som utvärderats inför studien (här utesluts de som inte uppfyllde inklusionskriterierna eller inte vill delta), antal som randomiseras, antal som allokerats till respektive interventionsarm och antal med slutförd behandlingen, samt antal och anledningar till avhopp vid varje steg.

Exponering (intervention, riskfaktor, fall): Motsvaras av faktorn man vill studera effekten av till exempel en intervention i en klinisk prövning (ex cervixcerklage), eller ett tillstånd/sjukdom/diagnos studiedeltagaren har (ex uterusanomali), eller ett beteende/vana (ex rökning) som medför att personen utsätts för en eventuell risk.

- Är det tydligt vad som klassas som exponerad? Vet vi något om följsamheten till behandlingen? Ex: Är exponeringen från en medicin baserat på uthämtade recept? Vet vi hur många som faktiskt tog medicinen? Finns det exponerade i den oexponerade gruppen individer som kan ha köpt medicin "over the counter"? Är diagnoskoderna som inhämtats från register korrekt satta/validerade?

Referensgrupp (oexponerad, jämförelsegrupp, kontroll): Motsvaras av de som inte är exponerade i en observationsstudie, ex de utan riskfaktorn i en kohortstudie, eller de utvalda kontrollerna vid en fallkontrollstudie eller interventionsstudie. Det är viktigt att kontrollerna liknar de exponerade i alla avseenden utom graden av exponering. Tips är att titta efter "Tabell 1" i en artikel där studiepopulationens karakteristika oftast beskrivs och jämförs baserat avseende exponeringsgrad. Generellt gäller att kontroller ska väljas oberoende av exponeringsstatus. Detta uppnås ofta genom att kontroller väljs slumpmässigt ur källpopulationen.

- Är kontrollgruppen adekvat utvald och tydligt definierad?
- Var ifrån har kontrollerna hämtats? Den allmänna befolkningen utifrån register? Sjukhuskontroller? Annat sätt?

- Har man matchat fall och kontroller? Baserat på vilka faktorer i sådana fall (ofta kön, ålder, län, födelseår)? Ju fler matchade faktorer, desto mer lika fall och kontroller, men desto svårare att hitta tillräckligt många matchade fall-kontroll-grupper. Statistiskt sett förbättras precisionen endast marginellt med >4 kontroller/fall.

Utfall (outcome):

- Är utfallet (eller effektvariabeln) den bästa möjliga, dvs undersöker studien det vi verkligen är intresserade av eller har de använt sig av surrogatmått (ex laboratorievärden, fysiologiska mått istället för ett kliniskt utfall)? Är utfallet rätt klassificerat? Är diagnoskoderna som använts från ett register korrekta/validerade?
- Finns det ett förutbestämt primärt utfall, eller flera? Det händer att ett stort antal variabler redovisas i metod och resultat, där man sedan i diskussionen fokuserar på de signifikanta resultaten.
- När har man valt att mäta utfallet? Tiden för uppföljning bör vara tillräckligt lång för att utfallet ska ha kunnat ske.

Huvudresultat: Ett studieresultat kan "förklaras" av tre orsaker: 1) det finns ett sant samband; 2) det har uppstått av slumpen (random error); 3) det har uppstått pga systematiska fel (bias).

- Vilka är huvudresultaten? Besvarar resultaten frågeställningen/hypotesen? Vilka slutsatser dras från artikeln?
- Är resultaten statistiskt signifikanta och kliniskt relevanta?
 - För att kunna avgöra om ett samband är statistiskt signifikant måste signifikansnivån anges, sedvanligt satt till $p < 0,05$. Det innebär att om studien görs om under exakt samma betingelser kommer vi i 5 fall av 100 inte kunna förkasta nollhypotesen (dvs inget samband). P visar således sannolikheten att skillnaden i ett resultat uppstått av slumpen. Påverkas framförallt av storleken på stickprovet/den studerade populationen. En väldigt liten skillnad kan bli signifikant om stickprovet är tillräckligt stort.
 - Ex: 20 000 kvinnor med hypertoni genomför ett dagligt motionsprogram. Efter en månad har deras medelartärbloodtryck sjunkit med 2 mmHg jämfört med kontrollgruppen, vilket beräknas vara en statistisk signifikant skillnad, men troligen inte så stor klinisk relevans. I praktiken skulle det lika gärna kunna handla om mätfel.
 - Konfidensintervall ger samma information som p-värden avseende på hypotestestning, men visar även storleken på skillnaden mellan grupperna som kan ge en hänvisning om klinisk relevans. Det "sanna" värdet finns inom intervallet med 95% sannolikhet.
- Är kriterier för kausalitet uppfyllda? För att tala om kausalitet 1) ska orsaken föregå verkan; 2) ska det finnas ett statistiskt samband mellan orsak och verkan, och 3) om man ändrar orsaken (interventionen) så ändras verkan (utfallet). Vid observationsstudier kan det vara svårt att visa att en ändring i exponeringsgrad förändrar utfallet. Då brukar man istället diskutera hur starkt sambandet är, om det sambandet finns i alla miljöer oavsett var det studeras, om man kan finna en biologisk modell som kan förklara sambandet och om det finns ett dos-responssamband.

Intern validitet: Den interna validiteten handlar hur väl en studies resultat motsvarar verkligheten och inte är ett resultat av systematiska fel (systematic error) - bias. Hög intern validitet = få/inga systematiska fel. Systematiska fel kan uppstå pga fel i studiedesignen (vilka som är med i studien och jämförs), fel i datainsamlingen (hur vi klassificerar exponering och utfall) och fel i analys av data (vad vi justerar för).

Systematiska fel brukar delas in i selektionsfel (selection bias), informationsfel (misclassification) och förväxlingsfaktorfel (confounding).

- *Selektionsfel (selection bias):* uppstår när utvalda studiedeltagare systematiskt skiljer sig avseende bakgrundsfaktorer från andra i bakgrundspopulationen som ej blir utvalda till studien. Ex i den kliniska prövningen om inte alla studiedeltagare har samma chans att randomiseras till en intervention eller möjlighet att slutföra en klinisk prövning. Denna risk minimeras (ej eliminieras) genom fullgod blindning. Kan även uppstå vid en fall-kontrollstudie om kontrollerna är felaktigt utvalda. En sorts selektionsbias sk "healthy worker bias" kan uppstå om arbetande studiedeltagare, som troligen är mer friska, jämförs med mindre friska studiedeltagare (extern kontrollgrupp) såsom den allmänna populationen. Då avspeglar inte de oexponerades risk för utfallet risken i den generella populationen.

Selektionsfel kan också uppstå om bortfallet från en studie är relaterat till både exponeringen och utfallet (sk differential loss-to-follow up exempelvis vid RCT (ex fler i interventionsarmen hoppar av pga biverkningar), kohortstudier).

- **Informationsfel** (missclassification): uppstår när en exponering, utfall eller andra variabler i en modell inte är korrekt uppmätt/värderade och leder till feltolkning av resultat. Kan uppstå i den oblandade och bländade kliniska prövningen, men även i observationsstudier där ex information om en variabel är felaktig. Resultaten påverkas olika beroende på om felklassificeringen är slumpmässig (*sk non-differential missclassification*, vilket ofta leder till en utspädning av ett samband) eller systematisk (*sk differential missclassification*, vilket kan både dölja och skapa ett samband som egentligen inte finns). Ett exempel på differential missclassification är övervakningsfel (surveillance bias) som kan uppstå om studiedeltagare/patienter som för sammanhanget har en betydelsefull exponering kontrolleras oftare än kontrollerna. Övervakningsfel kan också uppstå om patienten själv "lättare hittar tillbaka" till sjukvården om nya symptom uppstår och har en lägre tröskel att söka vård och då få fler diagnoser.
- **Förväxlingsfaktorfel** (confounding): innebär en sammanblandning av effekter. Ett samband mellan en exponering och ett utfall uppstår och förklaras helt eller delvis av att exponeringen samvarierar med en annan exponering (confounder) som direkt eller indirekt orsakar utfallet. Ex: låg födelsevikt (exponering) ökar risken för astma (utfall) hos barnet. Mammans rökning samvarierar med både födelsevikt och barnets astmautveckling (confounder).

Confounding åtgärdas genom randomisering, restriktion (undersök ex endast rökare eller icke-rökare om rökning är en confounder), stratifiering (sambandet mellan exponeringen och utfallet studeras i olika "strata" med ett konstant värde på confoundern), matchning (matcha fall och kontroller baserat på confoundern) eller regressionsanalys (en matematisk modell som beskriver hur utfallet varierar som en funktion av en eller flera exponeringar, inklusive confoundern).

Blindning: Alla som är involverade i en studie kan potentiellt påverka resultaten om de vet hur behandlingsalternativen fördelats. Detta gäller såväl patienter som studiepersonal och de som bedömer kliniska utfall. Detta kan medföra flera olika former av bias såsom placebo-effekt, performance bias (en sorts informationsfel där mer fokus läggs på de med interventionen) och detection bias (systematisk skillnad i hur utfallen bestäms). En RCT kan vara öppen (ej bländad), enkelbländad (antingen deltagare eller utövare), dubbelbländad (både deltagare och utövare) eller trippelbländad (även resultatbearbetningen utförs innan prövningskoden avslöjas).

Extern validitet: Kan studiens resultat generaliseras till andra människor (andra åldrar, kön, komorbiditetsgrad mm.), "settings" och tidsperioder? Förutsätter en hög intern validitet.

Precision: vid hög precision finns ett litet slumpmässigt fel ("random error"). Random error eller statistisk variabilitet uppstår pga att vi mäter saker hos ett stickprov av individer, och stickprovet skiljer sig ibland från vad som verkligen händer i hela populationen. Till skillnad från "systematic error" kan man minska "random error" och således öka precisionen genom att göra studiepopulationen större. Precisionen kan kvantifieras med konfidensintervall och p-värden (hög precision = "tamt" konfidensintervall, lågt p-värde).

Sample size: Antalet studiedeltagare som behövs för att man ska kunna visa på en statistisk signifikant skillnad mellan två eller flera grupper. Beror på spridningen av effektmåttet, skillnaden i effekt mellan grupperna, typ I-felet och typ II-felet.

- Typ I fel (α): risken att hävda att en skillnad föreligger mellan grupper även om så inte är fallet. Sätts vanligen till 0,05 (5% - risken att se en skillnad mellan grupperna som beror på slumpen är således 1:20). Notera att 0,05 är en arbiträr siffra, och kan sättas till 0,01 eller 0,005.
- Typ II fel (β): risken att missa en sann skillnad mellan grupperna. Sätts vanligen till 0,20 (20%).

Power: Istället för att redovisa typ-II felet, brukar man redovisa studiens power vilket ges av förhållandet $1-\beta$ (sannolikheten för typ II-felet) och brukar således sättas till 80%.

En studies power beror på: den valda signifikansnivån (α), stickprovsstorleken, storleken på den effekt (klinisk relevant skillnad) man vill kunna påvisa samt variabiliteten i populationen (standardavvikelsen). Bör beräknas innan man påbörjar studien. En poweranalys som görs efter datainsamlande kallas post-hoc och ger alltid låg power vid icke-signifikanta resultat. Då är det bättre att använda konfidensintervall för att se om skillnaden kan ha klinisk relevans. Pilotstudier kan användas för powerberäkning, likaså granskning av tillgänglig litteratur för förslag på effekt och varians.

Vill man visa att det inte finns någon statistisk signifikant skillnad mellan grupper (non-inferiority) krävs en power-beräkning innan, eftersom en avsaknad av statistisk signifikans i en studie inte betyder avsaknad av association, utan det kan vara fråga om en för liten studie.

Referenser:

Essential Medical Statistics. Betty R. Kirkwood, Jonathan A. C. Sterne. 2003. 2nd edition. Blackwell Science Ltd. Oxford, UK.

Att börja forska. Jonas F. Ludvigsson. 2015. 2a upplagan. Studentlitteratur. Lund, Sverige.

Gordis epidemiology. Celentano, David D. 2019. 6th edition. Elsevier. Philadelphia, USA.

Användbara checklistor:

CONSORT - <http://www.consort-statement.org/>: rapportering av kliniska randomiserade prövningar

STROBE - <https://www.strobe-statement.org/>: rapportering av observationsstudier.

PRISMA - <http://prisma-statement.org/PRISMAStatement/Checklist>: rapportering vid systematiska översikter

En hemsida som kan användas för att beräkna sample size: <https://clincalc.com/stats/samplesize.aspx>